基于多维泰勒网的多入多出非线性时滞系统辨识 *

李晨龙 [1, 2, 3, 4]，严洪森 [1, 2†]

(1. 东南大学自动化学院，南京 210096; 2. 复杂工程系统测量与控制教育部重点实验室，南京 210096; 3. 北京航空航天大学杭州创新研究院，杭州 310051; 4. 北京航空航天大学，北京 100191)

摘　要：针对多入多出 (Multiple Input Multiple Output，MIMO) 非线性时滞系统辨识的准确性与实时性问题，提出基于多维泰勒网 (Multi-dimensional Taylor Network，MTN) 的辨识方案。MTN 作为辨识模型，综合利用权剪枝 (Weight-Elimination，WE) 算法和共轭梯度算法 (Conjugate Gradient，CG)，即 WE-CG 算法作为 MTN 辨识模型的学习算法；WE 算法可以有效精简 MTN 辨识模型结构，从而降低计算复杂度，提高模型实时性能。最后，引入一个数值仿真例子和一个工程实例来验证所提辨识方案的有效性，同时与传统的 MTN 辨识方案作对比，给出了二者的准确性与复杂度分析，突出所提辨识方案的准确性与实时性。实验结果表明，所提方案够准确地对 MIMO 非线性时滞系统进行辨识。同时，相比传统的 MTN 辨识方案，所提辨识方案结构更精简，具有更低的算法复杂度。

关键词：多维泰勒网；辨识模型；多入多出系统；非线性时滞系统；剪枝算法

中图分类号：TP271　　doi: 10.19734/j.issn.1001-3695.2022.01.0031

Identification of MIMO nonlinear time-delay system using multi-dimensional Taylor network

Li Chenlong[1, 2, 3, 4], Yan Hongsen[1, 2†]

(1. *School of Automation*, *Southeast University*, *Nanjing* 210096, *China*; 2. *Key Laboratory of Measurement & Control of Complex Systems of Engineering*, *Ministry of Education*, *Southeast University*, *Nanjing* 210096, *China*; 3. *Hangzhou Innovation Institute*, *Beihang University*, *Hangzhou* 310051, *China*; 4. *Beihang University*, *Beijing* 100191, *China*)

Abstract: For the accuracy and real-time performance of multiple input multiple output (MIMO) nonlinear time-delay system, this paper propose the identification scheme based on multi-dimensional Taylor Network (MTN) . It use MTN as an identification model. MTN identification model's learning algorithm adopts the WE-CG algorithm, which combines the weight-elimination (WE) algorithm with the conjugate gradient (CG) algorithm. WE algorithm can simplify the structure of MTN model so as to reduce computational complexity and improve the real-time performance. Finally, there are two experimental examples containing a numerical example and a project one to verify the effectiveness of the proposed identification scheme. And use the traditional MTN identification scheme to compare. Both the accuracy and the computational complexity analysis illustrate the accuracy and the real-time performance of the proposed scheme. Results from the experiments and comparison show that the proposed scheme can identify the MIMO nonlinear time-delay system accurately, and the proposed identification scheme has a simpler structure and lower computational complexity than the traditional MTN identification scheme.

Key words: multi-dimensional Taylor network; identification model; MIMO system; nonlinear time-delay system; weight-elimination algorithm

0　引言

工业系统中大都具有非线性、时滞等特性，给建模和控制带来挑战[1,2]。以往在线性系统中，因为可以得到其精确的数学模型，所以利用传统的辨识方法可以满足其要求，比如最小二乘法、阶跃响应法等[3,4]。而对于非线性时滞系统辨识问题，由于很难获得系统精确的数学模型，因此，以往针对线性系统的辨识方法很难达到理想的效果，从而给辨识带来困难[5]。随着神经网络的出现，由于其强大的非线性逼近能力，给非线性辨识提供了一种有效的解决途径[6~10]。比如，文献[6]针对非线性系统，提出基于人工神经网络与切比雪夫函数的时滞自适应辨识方案；文献[7]基于时滞神经网络与核函数的内部关系，提出针对工程领域的非线性系统辨识方案；文献[8]针对带有未知时滞的非线性连续系统，提出基于时滞神经网络的辨识方案；文献[9]针对非线性系统辨识，提出了一种新型神经网络方案；文献[10]针对非线性动态系统，提出了一种基于数据知识的模糊神经网络辨识方案。然而，神经网络由于在训练过程中花费时间长，易陷入极小值点，且算法复杂度较高，从而不能满足实时构建预测模型的要求[11]。此外，相比于单入单出(Single Input Single Output，SISO)系统，多入多出 (Multiple Input Multiple Output，MIMO) 系统由于输入和输出变量之间的关系更加复杂而更加难以建模[12~14]。

针对以上问题，严洪森教授于 2010 年 10 月提出多维泰勒网的建模思想[15,16]，之后被应用在非线性时间序列建模与故障监测上[17~21]。之后在 2010 年 12 月严洪森教授提出多维泰勒网优化控制的思想[15,16]，应用于非线性系统控制领域[22~24]，先后对非线性时变系统[22]、非线性时滞系统[23]与非线性随机系统[24]

收稿日期：2022-01-07；**修回日期：**2022-03-17　**基金项目：**国家自然科学基金资助项目(61673112，60934008)；中央高校基本科研业务费专项资金资助项目(2242017K10003，2242014K10031)；上海航天科技创新基金资助项目(SAST2019-020)；江苏高校优势学科建设工程资助项目；浙江省博士后科研项目择优资助基金资助项目(ZJ2021074)

作者简介：李晨龙(1988-)，男，辽宁沈阳人，助理研究员，博士，主要研究方向为系统辨识、预测控制、多维泰勒网优化控制等研究；严洪森(1957-)，男(通信作者)，浙江江山人，教授，博导，博士，主要研究方向为知识化制造、多维泰勒网优化控制等(hsyan@seu.edu.cn)．

的控制问题进行了广泛研究。多维泰勒网与多维泰勒网优化控制器具有以下特点[15~24]：a) MTN 具有可以任意逼近非线性函数的能力；b) MTN 模型与控制器，其结构简单，只包含加法和乘法，不含有超越函数和无理函数，参数调整方便，便于工程实现，无须复杂的数学计算，便于实现实时建模与控制；c) MTN 控制器本身具有“学习”能力，使它在控制过程中不断地完善自己，使控制效果越来越好，提高控制器的鲁棒性能；d) MTN 控制器具有良好的可扩展性，能够融合其他控制方法或者控制思想，具有一定的容错性和进化能力。

在基于此，本文针对 MIMO 非线性时滞系统，提出基于 MTN 的辨识方案。首先，MTN 作为辨识模型；然后，综合利用权剪枝(Weight-Elimination, WE)算法和共轭梯度算法(Conjugate Gradient, CG)，即 WE-CG 算法作为 MTN 辨识模型的学习算法；WE 算法对 MTN 辨识模型结构进行精简从而降低计算复杂度以满足实时性要求；最后，引入一个数值仿真例子和一个工程实例-连续搅拌反映器(continuous stirred tank reactor, CSTR)来验证本文所提方案的有效性，同时引入传统的 MTN 辨识方案[17]作对比，给出了二者的准确性与复杂度分析。

本文的主要贡献如下：

a) 针对 MIMO 非线性时滞系统，借助 MTN 的强大逼近性能，提出基于 MTN 的辨识方案，所提方案易于实施；

b) WE-CG 算法作为 MTN 辨识模型的学习算法，WE 算法可以有效精简 MTN 辨识模型结构，从而降低算法复杂度，提高模型实时性能；

c) 所提方案具有较低的计算复杂度，即实时性能较好；

d) 所提辨识方案的具有一般性和实际应用价值。

1　系统描述与多维泰勒网

1.1　系统描述

MIMO 离散非线性时滞系统由如下表示：

$$\begin{aligned}\boldsymbol{y}(k)=\boldsymbol{g}[&\boldsymbol{y}(k-1),\boldsymbol{y}(k-2),\cdots,\boldsymbol{y}(k-n_y),\\&\boldsymbol{u}(k-d),\cdots,\boldsymbol{u}(k-n_u-d)]\end{aligned}\tag{1}$$

其中，$\boldsymbol{g}(\cdot)=[g_1(\cdot),\cdots,g_q(\cdot)]^{\mathrm{T}}\in R^q$ 为未知的非线性向量函数；$\boldsymbol{y}(k)=[y_1(k),\ldots,y_q(k)]^{\mathrm{T}}\in \mathrm{R}^q$；$\boldsymbol{u}(k)=[u_1(k),\ldots,u_p(k)]^{\mathrm{T}}\in \mathrm{R}^p$ 分别为系统(1)的 q 维输出和 p 维输入；n_{y_j} （$j=1,2,\ldots,q$），n_{u_i} （$i=1,2,\ldots,p$）分别为第 j 个输出分量和第 i 个输出分量的阶次，且 $n_y=\{n_{y_1},n_{y_2},\ldots,n_{y_q}\}$，$n_u=\{n_{u_1},n_{u_2},\ldots,n_{u_p}\}$；$d$ 为系统的时滞。

则，在 k 时刻，系统(1)的第 j 个输出分量可以表示为

$$\begin{aligned}y_j(k)=g_j(&y_1(k-1),\cdots,y_1(k-n_{y_1}),\cdots,\\&y_q(k-1),\cdots,y_q(k-n_{y_q}),\\&u_1(k-d),u_1(k-d-1),\cdots,\\&u_1(k-d-n_{u_1}),\cdots,u_p(k-d),\\&u_p(k-d-1),\cdots,u_p(k-d-n_{u_p}))\end{aligned}\tag{2}$$

1.2　多维泰勒网

1.2.1 MIMO 多维泰勒网

针对于 MIMO 非线性系统，基于多维泰勒网，第 j 个输出分量 $\hat{y}_j(k+1)$ 可写成如下形式[15]

$$\hat{y}_j(k+1)=\sum_{\hat{r}=1}^{\hat{N}(\hat{\imath},\hat{m})}\hat{w}_{j,\hat{r}}(k)\prod_{\hat{s}=1}^{\hat{\imath}}\hat{z}_{\hat{s}}^{\hat{\lambda}(\hat{r},\hat{s})}(k)\tag{3}$$

其中，$\hat{N}(\hat{\imath},\hat{m})$ 表示 $\hat{\imath}$ 元函数 $\hat{f}_j(\cdot)$ 展开成 $\hat{m}$ 次逼近多项式乘积项的总项数，权重向量为 $\hat{\boldsymbol{w}}_j(k)=[\hat{w}_{j,1}(k),\cdots,\hat{w}_{j,\hat{N}(\hat{\imath},\hat{m})}(k)]^{\mathrm{T}}$，$\hat{w}_{j,\hat{r}}(k)$ 表示式(3)中第 $\hat{r}$ 个乘积项的权重系数，$\hat{\lambda}(\hat{r},\hat{s})$ 表示第 $\hat{r}$ 个乘积项中变量 $\hat{z}_{\hat{s}}(k)$ 的次数，且 $\sum_{\hat{s}=1}^{\hat{\imath}}\hat{\lambda}(\hat{r},\hat{s})\le\hat{m}$。

MIMO 非线性系统的 MTN 结构图如图 1 所示。

1.2.2 多维泰勒网逼近性能

引理 1[25]　任何定义于一个闭区间的连续函数可以用多项式函数任意准确地逼近。

引理 2[18]　对于定义于一个闭区间的连续函数 $f(x_1,x_2,\ldots,x_n)$，可以用 $\sum_{t=1}^{N(n,m)}w_t\prod_{i=1}^{n}x_i^{\lambda_{t,i}}$ 逼近。其中 $N(t,m)$ 为逼近展开式中乘积项的总项数，w_t 为逼近展开式中第 t 个乘积项之前的权值，$\lambda_{t,i}$ 为展开式中第 t 个乘积项中变量 x_i 的幂次。

正如图 1 所示，MTN 采用前向单中间层结构，包括输入层、中间层和输出层。同时基于引理 1 和引理 2 说明，只要 $N(t,m)$ 足够大，它就能够以足够的精度逼近任意模型[18,25]。

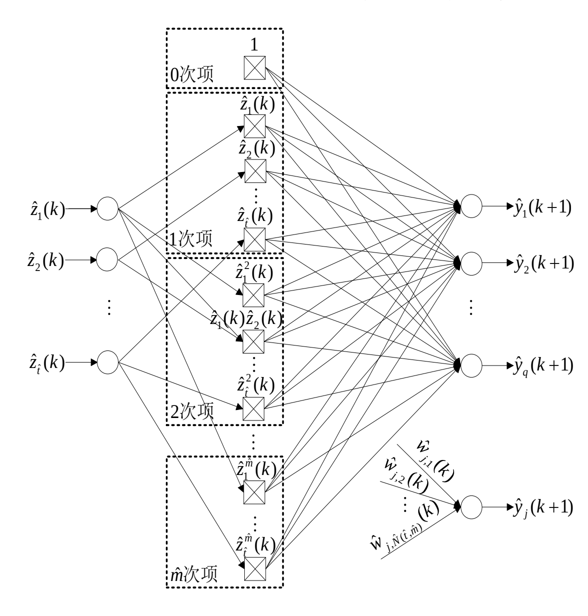


图 1　MIMO 的 MTN 结构图

Fig. 1　MIMO MTN structure

1.2.3 多维泰勒网复杂度分析

由于使用相同的计算机，相同的时间来做实时性的分析是不可能的，因此本文引用浮点数来估计方案的计算量，即实时性[26,27]。一个加号和一个乘号都叫做一个浮点，加号和乘号的总数作为总的浮点数，即浮点数的总和代表着计算复杂度。当展开次幂为 2 次，表 1 为 MTN 在一次迭代中的计算复杂度[17]，突出 MTN 因其结构简单而带来出色的实时性能。

表 1　MTN 的计算复杂度

Tab. 1　Computation complexity of MTN model

操作项	MTN
中间节点数	$n_0(n_0-1)/2+2n_0+1$
权值	$n_1(n_0(n_0-1)/2+2n_0+1)$
加法	$n_1(n_0(n_0-1)/2+2n_0)$
乘法	$n_1(n_0(n_0-1)/2+2n_0+1)$
$f(\cdot)$	n_1

表中 n_0 和 n_1 分别表示 MTN 输入和输出层的节点数。MTN 只包括加法与乘法，其计算复杂度仅相当于神经网络中单个神经元的泰勒展开。

2　MIMO 非线性时滞系统多维泰勒网辨识方案

本文采用 MTN 作为辨识模型，WE-CG 算法作为学习算法，WE 算法可以有效精简 MTN 网络结构。

2.1　MTN 辨识模型

对于系统(1)，$\boldsymbol{g}[\cdot]$ 可以由 MTN 以任意精度逼近，其映射关系记为 $\hat{\boldsymbol{g}}[\cdot]$，则 MTN 辨识模型可以写成如下形式：

$$\begin{aligned}\hat{\boldsymbol{y}}(k)=\hat{\boldsymbol{g}}_1[&\boldsymbol{y}(k-1),\boldsymbol{y}(k-2),\cdots,\boldsymbol{y}(k-n_y),\\&\boldsymbol{u}(k-d),\cdots,\boldsymbol{u}(k-n_u-d),\hat{\boldsymbol{w}}_1(k)]\end{aligned}\tag{4}$$

其中，$\hat{\boldsymbol{y}}(k)$ 为 MTN 辨识模型输出，且 $\hat{\boldsymbol{y}}(k)=[\hat{y}_1(k),\cdots,\hat{y}_q(k)]^{\mathrm{T}}\in R^q$；

$\hat{w}_1(k)$ 为 MTN 权系数向量；$\hat{g}_1(\cdot)=[\hat{g}_1(\cdot),\cdots,\hat{g}_q(\cdot)]^T\in R^q$；$\boldsymbol{u}(k)$，$\boldsymbol{y}(k)$，$n_y$，$n_u$ 和 d 与式(1)有相同的含义。

然后，系统(4)在第 k 时刻的第 j 次输出分量为

$$\begin{aligned}\hat{y}_j(k)=\hat{g}_j[&y_1(k-1),\cdots,y_1(k-n_{y_1}),\cdots,y_q(k-1),\cdots,\\&y_q(k-n_{y_q}),u_1(k-d),u_1(k-d-1),\cdots,\\&u_1(k-d-n_{u_1}),\cdots,u_p(k-d),u_p(k-d-1),\cdots,\\&u_p(k-d-n_{u_p}),w_j]\end{aligned}\tag{5}$$

为了方便而不失一般性，记

$$\hat{t}_u=\sum_{i=1}^{p}(n_{u_i}+1)$$

$$\hat{t}_y=\sum_{j=1}^{q}(n_{y_j}+1)$$

$$\hat{t}=\hat{t}_u+\hat{t}_y$$

$$\begin{aligned}\hat{\boldsymbol{z}}(k)&=[\hat{z}_1(k),\cdots\hat{z}_{\hat{t}_y}(k),\hat{z}_{\hat{t}_y+1}(k),\cdots,\hat{z}_{\hat{t}}(k)]^T=\\&[y_1(k\text{-}1),\cdots,y_q(k-n_{y_q}),u_1(k\text{-}d),\cdots,u_p(k-n_{u_p}-d)]^T\end{aligned}$$

MTN 的第 j 次输出分量 $\hat{y}_j(k)$ 的权值向量为

$$\hat{\boldsymbol{w}}_j(k)=[\hat{w}_{j,1}(k),\ldots,\hat{w}_{j,\hat{N}(\hat{t},\hat{m})}(k)]^T$$

则，相应的 MTN 模型(5)可以改写成如下形式[17]

$$\hat{y}_j(k)=\sum_{\hat{r}=1}^{\hat{N}(\hat{t},\hat{m})}\hat{w}_{j,\hat{r}}(k)\prod_{\hat{s}=1}^{\hat{t}}\hat{z}_{\hat{s}}^{\hat{\lambda}(\hat{r},\hat{s})}(k)\tag{6}$$

其中，$\hat{N}(\hat{t},\hat{m})$ 表示 $\hat{t}$ 元函数 $\hat{g}_j(\cdot)$ 展开成 $\hat{m}$ 次逼近多项式乘积项的总项数，$\hat{w}_{j,\hat{r}}(k)$ 表示式(6)中第 $\hat{r}$ 个乘积项的权重系数，$\hat{\lambda}(\hat{r},\hat{s})$ 表示第 $\hat{r}$ 个乘积项中变量 $\hat{z}_{\hat{s}}(k)$ 的次数，且 $\sum_{\hat{s}=1}^{\hat{t}}\hat{\lambda}(\hat{r},\hat{s})\le\hat{m}$。

2.2　模型学习算法

2.2.1 共轭梯度法

MTN 参数可以通过采样学习与适当的学习算法来获得。

设采样集 $T=\{(\boldsymbol{u}(k),\boldsymbol{y}(k+1))\}_{k=0}^{N-1}$，$\boldsymbol{u}(k)\in R^p$ 为系统输入，$\boldsymbol{y}(k+1)\in R^q$ 为系统输出。误差变化方程如下：

$$\begin{aligned}\hat{E}&=\frac{1}{N}\sum_{k=0}^{N-1}(\boldsymbol{y}(k+1)-\hat{\boldsymbol{y}}(k+1))^2\\&=\frac{1}{N}\sum_{k=1}^{N-1}\Big(y_j(k+1)-\sum_{\hat{r}=1}^{\hat{N}(\hat{t},\hat{m})}\hat{w}_{j,\hat{r}}(k)\prod_{\hat{s}=1}^{\hat{t}}\hat{z}_{\hat{s}}^{\hat{\lambda}(\hat{r},\hat{s})}(k)\Big)^2\end{aligned}\tag{7}$$

设 $\quad\hat{\boldsymbol{c}}(k)=\left(\prod_{\hat{s}=1}^{\hat{t}}\hat{z}_{\hat{s}}^{\hat{\lambda}(1,\hat{s})}(k),\prod_{\hat{s}=1}^{\hat{t}}\hat{z}_{\hat{s}}^{\hat{\lambda}(2,\hat{s})}(k),\ldots,\prod_{\hat{s}=1}^{\hat{t}}\hat{z}_{\hat{s}}^{\hat{\lambda}(\hat{N}(\hat{t},\hat{m}),\hat{s})}(k)\right)^T$

$$\hat{\boldsymbol{A}}=(\hat{c}(0),\hat{c}(1),\ldots,\hat{c}(N-1))$$

$$\hat{\boldsymbol{w}}_j(k)=\left(\hat{w}_{j,1}(k),\hat{w}_{j,2}(k),\cdots,\hat{w}_{j,\hat{N}(\hat{t},\hat{m})}(k)\right)^T,\ j=1,2,\ldots,q$$

$$\hat{\boldsymbol{w}}_1=(\hat{w}_1(k),\hat{w}_2(k),\ldots,\hat{w}_{\hat{N}(\hat{t},\hat{m})}(k))^T$$

则，式(7)可以写成如下形式：

$$\begin{aligned}\hat{E}&=\frac{1}{N}\sum_{k=0}^{N-1}(\hat{\boldsymbol{w}}_1^T\hat{\boldsymbol{c}}(k)-\boldsymbol{y}(k+1))^2=\\&\frac{1}{N}\hat{\boldsymbol{w}}_1^T\sum_{k=0}^{N-1}(\hat{\boldsymbol{c}}(k)\hat{\boldsymbol{c}}^T(k))\hat{\boldsymbol{w}}_1-\\&\frac{2}{N}\Big(\sum_{k=0}^{N-1}\boldsymbol{y}(k+1)\hat{\boldsymbol{c}}^T(k)\Big)\hat{\boldsymbol{w}}_1+\frac{1}{N}\sum_{k=0}^{N-1}y^2(k+1)\end{aligned}\tag{8}$$

计算 $\hat{E}$ 关于 $\hat{\boldsymbol{w}}_1$ 的偏导数，得

$$\frac{\partial\hat{E}}{\partial\hat{\boldsymbol{w}}_1}=\frac{2}{N}\sum_{k=0}^{N-1}(\hat{\boldsymbol{c}}(k)\hat{\boldsymbol{c}}^T(k))\hat{\boldsymbol{w}}_1-\frac{2}{N}\sum_{k=0}^{N-1}\boldsymbol{y}(k+1)\hat{\boldsymbol{c}}(k)\tag{9}$$

设 $\quad\hat{\boldsymbol{g}}=\dfrac{\partial\hat{E}}{\partial\hat{\boldsymbol{w}}_1}$，$\hat{\boldsymbol{Q}}=\dfrac{2}{N}\sum_{k=0}^{N-1}(\hat{\boldsymbol{c}}(k)\hat{\boldsymbol{c}}^T(k))$，$\hat{\boldsymbol{l}}=-\dfrac{2}{N}\sum_{k=0}^{N-1}\boldsymbol{y}(k+1)\hat{\boldsymbol{c}}(k)$

则，可以得到

$$\hat{\boldsymbol{g}}=\hat{\boldsymbol{Q}}\hat{\boldsymbol{w}}_1+\hat{\boldsymbol{l}}\tag{10}$$

设 $\boldsymbol{y}=(\boldsymbol{y}(1),\boldsymbol{y}(2),\cdots,\boldsymbol{y}(N))^T$，$\hat{\boldsymbol{Q}}$，$\hat{\boldsymbol{l}}$ 和 $\hat{\boldsymbol{g}}$ 可以重新写成如下形式：

$$\hat{\boldsymbol{Q}}=\frac{2}{N}\hat{\boldsymbol{A}}\hat{\boldsymbol{A}}^T,\quad\hat{\boldsymbol{l}}=-\frac{2}{N}\hat{\boldsymbol{A}}\boldsymbol{y},\quad\hat{\boldsymbol{g}}=\frac{2}{N}\hat{\boldsymbol{A}}(\hat{\boldsymbol{A}}^T\hat{\boldsymbol{w}}_1-\boldsymbol{y})。$$

为了获得一个精确的模型，权重系数向量 $\hat{\boldsymbol{w}}_1$ 的值应该通过数据 $\{\boldsymbol{u}(k),\boldsymbol{y}(k+1)\}$ 来更新。梯度法是一种常用的优化方法，其可以对参数进行调整，即，$\hat{\boldsymbol{w}}_1$ 可以根据每次离线 MTN 的负梯度方向来更新一次。设 $\hat{\boldsymbol{w}}_\tau$ 为 $\hat{\boldsymbol{w}}_1$ 经过 τ 步训练后的值，则 $\hat{\boldsymbol{w}}_{\tau+1}=\hat{\boldsymbol{w}}_\tau-\hat{\boldsymbol{\mu}}_\tau\hat{\boldsymbol{g}}_\tau$，其中 $\boldsymbol{\mu}_\tau=\dfrac{\hat{\boldsymbol{g}}_\tau^T\hat{\boldsymbol{g}}_\tau}{\hat{\boldsymbol{g}}_\tau^T\hat{\boldsymbol{Q}}\hat{\boldsymbol{g}}_\tau}$。然而，梯度法采用锯齿形的方式到达最小值点，所以其搜索方向会一直保持垂直。幸运的是，采用共轭梯度法可以有效地解决这一问题，其权重可以更新如下：

$$\hat{\boldsymbol{w}}_{\tau+1}=\hat{\boldsymbol{w}}_\tau+\boldsymbol{\mu}_\tau\hat{\boldsymbol{p}}_\tau\tag{11}$$

其中，$\hat{\boldsymbol{p}}_\tau=-\hat{\boldsymbol{g}}_\tau+\hat{\boldsymbol{s}}_{\tau-1}\hat{\boldsymbol{p}}_{\tau-1}$，$\boldsymbol{\mu}_\tau=\dfrac{\hat{\boldsymbol{g}}_\tau^T\hat{\boldsymbol{g}}_\tau}{\hat{\boldsymbol{g}}_\tau^T\hat{\boldsymbol{Q}}\hat{\boldsymbol{g}}_\tau}$，$\hat{\boldsymbol{s}}_{\tau-1}=\dfrac{\hat{\boldsymbol{g}}_\tau^T\hat{\boldsymbol{Q}}\hat{\boldsymbol{p}}_{\tau-1}}{\hat{\boldsymbol{p}}_{\tau-1}^T\hat{\boldsymbol{Q}}\hat{\boldsymbol{p}}_{\tau-1}}$，初始值 $\hat{\boldsymbol{p}}_1=-\hat{\boldsymbol{g}}_1$。

2.2.2 WE-CG 算法

随着 MTN 的输入节点增加，会导致“维数爆炸”，因此有必要精简网络结构。权剪枝算法可用于精简 MTN 的中间节点，以此来减小网络的复杂度。本节通过结合 CG 法与 WE 法，即 WE-CG 法作为 MTN 辨识模型的学习算法。

引理 3[28]　假设目标函数 $h(x)$ 可由最小结构网络 $Net_{(h-1)}$ 实现，网络 $Net_{(h)}$ 比 $Net_{(h-1)}$ 多一个中间节点。如果 $Net_{(h)}$ 能实现 $h(x)$，则 $Net_{(h)}$ 或者有 $h-1$ 个中间节点与中间节点 $Net_{(h-1)}$ 完全重合，而余下的那个独立中间节点输出权值为零；或者 $Net_{(h)}$ 中有 $h-2$ 个中间节点与 $Net_{(h-1)}$ 中的中间节点完全重合，而余下的两个重合的中间节点叠加后与 $Net_{(h-1)}$ 中最后一个中间节点相等。

根据引理 3，当训练样本充分多时，中间节点的衰减和重合是必然的。因此，有必要使用剪枝算法来精简网络，在去掉冗余节点后的网络仍然可以表示一个目标函数。本文通过正则化实现对 MTN 冗余节点删除来精简 MTN 网络结构。

由于该正则化项具有剪枝特性[29,30]，因此一些冗余的输出权值将逐步衰减到零，从而可以删除与之相连的中间节点，基于此，本节在 CG 算法的基础上引入带有高斯正则化项的权剪枝算法作为 MTN 辨识模型的学习算法。

本文采取如下目标函数：

$$J(\boldsymbol{w}_1)=\hat{E}(\boldsymbol{w}_1)+\gamma C(\boldsymbol{w}_1)\tag{12}$$

其中，$\hat{E}(\boldsymbol{w}_1)$ 通常为网络误差平方和；等式右边第二项为正则化项；γ 为正则化系数，且 $\gamma\ge0$；$C=-\dfrac{2w_i^2/w_0^2}{(1+w_i^2/w_0^2)^2}$，且 w_0 为基准权值。当某些中间节点的输出权值满足式(13)时，去掉中间节点。

$$|w_i|<w_{\min}\tag{13}$$

其中，$w_{\min}$ 为临界权值。由于正则化系数 γ 的取值对学习结果有很大影响，因此提出一种在学习过程中动态改变 γ 的方案，即在每次数据中心修正后，动态修改 γ。MTN 辨识模型 $\boldsymbol{w}_1$ 的调整方法采用如下等式[29,30]。

$$\boldsymbol{w}_1(k+1)=\boldsymbol{w}_1(k)+\hat{\mu}_\tau\hat{\boldsymbol{p}}_\tau+\gamma C(\boldsymbol{w})\tag{14}$$

其中，$\hat{\mu}_\tau$，$\hat{\boldsymbol{p}}_\tau$ 与式(11)中的含义相同；γ，$C(\boldsymbol{w})$ 与式(12)中的含义相同。

本节参照 Weigend 等人[29,30]给出的正则化参数调整方法，在学习过程中随时检测以下误差量之间的关系：

$E(k)$：当前时刻权值调节时的误差；

$E(k-1)$：前一时刻权值调节时的误差；

$A(k)$：当前时刻的加权平均误差，定义为 $A(k)=\kappa A(k-1)+(1-\kappa)E(k)$，其中 κ 为接近于 1 的系数；

D：期望误差值，如果没有先验知识，可设定 $D=0$，但此时计算时间可能较长。

对 γ 进行动态调节，具体规则如下：

如果 $E(k)<E(k-1)$，或 $E(k)<D$，则 $\gamma(k)=\gamma(k-1)+\Delta\gamma$；

如果 $E(k)\ge E(k-1)$，或 $E(k)<A(k)$，且 $E(k)\ge D$，则 $\gamma(k)=\gamma(k-1)-\Delta\gamma$；

如果 $E(k)\ge E(k-1)$，或 $E(k)\ge A(k)$，且 $E(k)\ge D$，则 $\gamma(k)=\sigma\gamma(k-1)$，其中 σ 为接近 1 的系数。

2.3 MTN 辨识方案

本文针对 MIMO 非线性时滞系统，提出基于 MTN 的辨识方案。首先，MTN 作为辨识模型；其次，WE-CG 算法作为 MTN 辨识模型的学习算法；WE 算法对 MTN 辨识模型结构进行精简从而降低计算复杂度以满足实时性要求。具体过程如算法 1 所示。

算法 1 基于 MTN 的 MIMO 非线性时滞系统辨识

步骤 1:确定 MTN 辨识模型结构与初始权值；

步骤 2:按照误差标准函数(7)，MTN 辨识模型由 WE-CG 算法进行训练；

步骤 3:由式(14)更新 $\hat{w}_1$；

步骤 4:重复步骤 2-4，直到误差满足给定精度停止训练。

3 实验结果及分析

本文通过一个数值仿真例子和一个实例-连续搅拌反映器系统(continuous stirred tank reactor, CSTR)过程，来验证所提方案的准确性与实时性。

3.1 算例

考虑如下的 MIMO 非线性时滞系统：

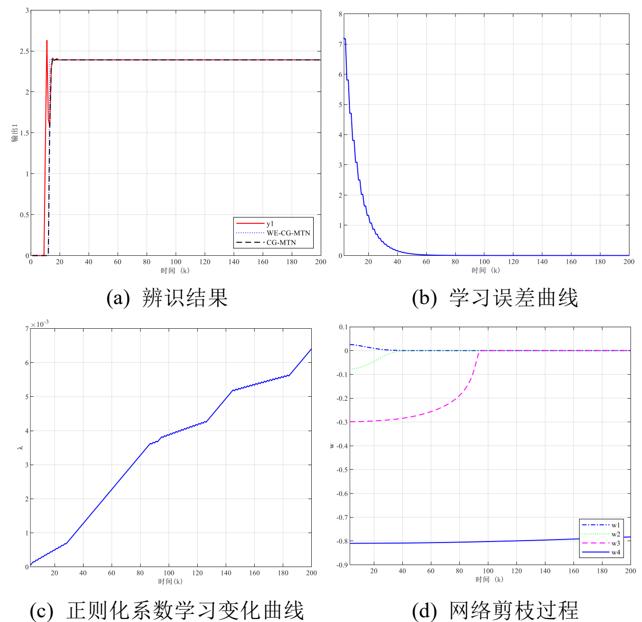
$$y_1(k)=\frac{2y_1(k-1)u_1^2(k-1)}{1+y_1^2(k-1)}+0.3u_1(k-1-d)+0.7u_1(k-2)+0.2u_2(k-d)+\sin(0.5u_2(k-d))$$

$$y_2(k)=\frac{1.7y_2(k-1)u_2^2(k-1)}{1+y_2^2(k-1)}+0.5u_2(k-1-d)+0.5u_2(k-2)+0.2u_1(k-d)+\sin(0.5u_2(k-d-2))$$

其中，$d=6$。

输入为单位阶跃信号，MTN 辨识模型采用 6-28-2 的结构，即 6 个输入节点，展开到 2 次，WE-CG 算法作为其学习算法。CG 学习算法设定如下：迭代次数为 200。WE 算法设定如下：参考权值为 0.1，正则化系数初值为 0，滤波系数为 κ=0.092，γ=0.995，正则化系数的增量为 0.5×10^{-4}，消除冗余权重的阈值为 0.05。其中用 200 个采样时刻进行训练，100 个采样时刻用于测试。

在图 2、3 中，y1 和 y2 为系统输出，WE-CG-MTN 为所提辨识方案，CG-MTN 为对比方案。

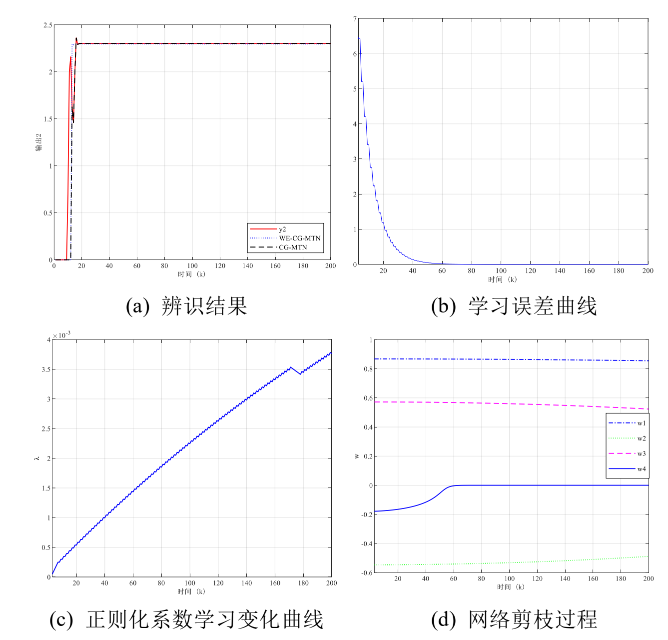


(a) 辨识结果 (b) 学习误差曲线

(c) 正则化系数学习变化曲线 (d) 网络剪枝过程

图 2 子系统 1 辨识

Fig. 2 Identification of sub-system 1



(a) 辨识结果 (b) 学习误差曲线

(c) 正则化系数学习变化曲线 (d) 网络剪枝过程

图 3 子系统 2 辨识

Fig. 3 Identification of sub-system 2

图 2 与 3 为辨识过程，其中，图 2(a)和图 3(a)为辨识结果，图 2(b)和图 3(b)为学习误差曲线，图 2(c)和图 3(c)为正则化系数的学习变化曲线，图 2(d)和图 3(d)为网络剪枝过程。尽管正则化系数的变化并不是平滑的学习过程，但 WE 算法的平滑的学习曲线表明 MTN 权重变化过程不会导致严重波动的训练误差，并对最后的剪枝的结果影响较小。

从仿真结果中本文可以看到，针对第一个子系统，WE-CG-MTN 方案和 CG-MTN 方案训练的均方误差分别为 0.0318 和 0.0285，测试的均方误差分别为 0.0365 和 0.0364；对于第二个子系统，WE-CG-MTN 方案和 CG-MTN 方案训练的均方误差分别为 0.0172 和 0.0172，测试的均方误差分别为 0.0521 和 0.0520。可以看出，WE 算法的引入虽然对学习误差有少许影响，但是对于两个子系统，去除冗余中间节点后，其节点分别只有 11 个和 19 个，这大大减少了计算量，增加了实时性能。WE-CG-MTN 与 CG-MTN 辨识方案的复杂度分析如表 2 所示。

表 2 复杂度分析

Tab. 2 Computation complexity analysis

预测模型构建方案	结构	加法	乘法	浮点数总和
WE-CG-MTN	6-11-2	20	22	42
CG-MTN	6-28-2	54	56	110

从表 2 看到经过精简后的 MTN 辨识模型的浮点数远远小于经典的 MTN 辨识方案，也就是说本文的计算量远远小于经典的 MTN 辨识方案，即所提方案具有很好的实时性能。

需要指出的是，MTN 相当于神经网络的一个神经元，其计算量远远小于神经网络，相关文献已经对传统 MTN 辨识模型和 BP 神经网络辨识模型的准确性与实时性进行了对比分析，表明传统 MTN 辨识模型优于 BP 神经网络辨识模型[19]。这里，本文对比了所提方案与传统 MTN 辨识方案的准确性与实时性分析，通过图 2、3 与表 2 看出，从准确性分析角度，无论是训练的均方误差，还是测试的均方误差，本文所提辨识方案与传统 MTN 辨识方案几乎属于同一数量级，而从计算复杂度角度，本文所提辨识方案的计算量少于传统的 MTN 辨识方案，即实时性更优。

3.2 实例-CSTR 过程

CSTR 为厌氧处理过程，其为一个由原油与微生物完全混合的状态[31~34]，如图 4 所示。CSTR 生成装置是带有搅拌装置的密闭水箱，其是一个具有时滞的典型非线性化工过程。

CSTR 模型采用如下模型[35]：

$$\frac{dx_1(t)}{dt}=-x_1(t)+D_a[1-x_1(t)]\exp(\frac{x_2(t)}{1+x_2(t)/\varphi})+C[u_1(t-\tau)-x_1(t)]$$

$$\frac{dx_2(t)}{dt}=-x_2(t)+BD_a[1-x_1(t)]\exp(\frac{x_2(t)}{1+x_2(t)/\varphi})+C[u_2(t-\tau)-x_2(t)]$$

离散化，得

$$x_1(k)=x_1(k-1)-x_1(k-1)\Delta T+D_a[1-x_1(k-1)]\exp(\frac{x_2(k-1)}{1+x_2(k-1)/\varphi})\Delta T+C[u_1(k-d)-x_1(k)]\Delta T$$

$$x_2(k)=x_2(k-1)-x_2(k-1)\Delta T+BD_a[1-x_1(k-1)]\exp(\frac{x_2(k-1)}{1+x_2(k-1)/\varphi})\Delta T+C[u_2(k-d)-x_2(k)]\Delta T$$

其中，$D_a=0.008$，$\varphi=10$，$B=6$，$C=0.5$，$d=6$，采样时间 ΔT 为 0.1s。

$$y_1(k)=x_1(k)\text{；}\quad y_2(k)=x_2(k)\text{ 。}$$

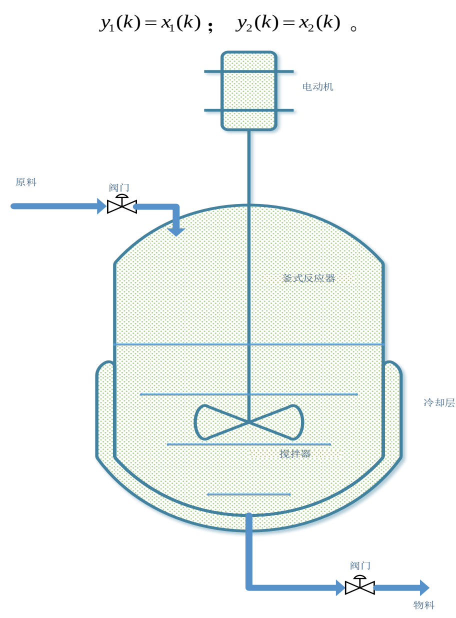


图 4　CSTR 过程

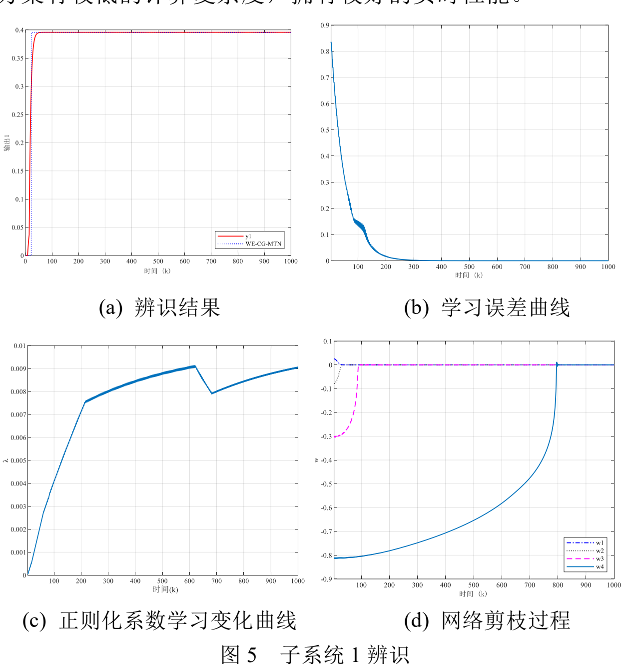
Fig. 4　CSTR process

输入采用单位阶跃信号，MTN 辨识模型采用 6-28-2 的结构，即 6 个输入节点，展开到 2 次，WE-CG 算法作为其学习算法。CG 学习算法设定如下：迭代次数为 1000。WE 算法设定如下：参考权值为 0.1，正则化系数初值为 0，滤波系数为 κ=0.092，γ=0.995，正则化系数的增量为 0.5×10^{-4}，消除冗余权重的阈值为 0.05。其中用 1000 个采样时刻进行训练，200 个采样时刻用于测试。

图 5 与图 6 为辨识过程，其中，图 5(a)和图 6(a)为辨识结果，图 5(b)和图 6(b)为学习误差曲线，图 5(c)和图 6(c)为正则化系数的学习变化曲线，图 5(d)和图 6(d)为网络剪枝过程。尽管正则化系数的变化并不是平滑的学习过程，但 WE 算法的平滑的学习曲线表明 MTN 权重变化过程不会导致严重波动的训练误差，并对最后的剪枝的结果影响较小。

从仿真结果中本文可以看到，针对第一个子系统，WE-CG-MTN 方案训练的均方误差为 0.0024，测试的均方误差为 0.0052；对于第二个子系统，WE-CG-MTN 方案训练的均方误差为 0.0065，测试的均方误差分别为 0.0108。可以看出，WE 算法的引入虽然对学习误差有少许影响，但是对于两个子系统，去除冗余中间节点后，其节点分别只有 2 个和 7 个，这大大减少了计算量，增加了实时性能。

需要说明的是 MTN 辨识模型结构的选择，本文采用 6-28-2 的结构，即 6 个输入节点，展开到 2 次，由图 1 可知中间层节点数为 28 个。当输入节点数不变，随着展开次数的增加，中间节点数会随之增加，造成节点冗余的发生。对于 WE 参数的选取，其决定了剪枝的效果，通常通过多次实验来确定参数的选择。本文看到在引入 WE 算法后，去除了中间冗余节点，减少了计算量。但是由于使用相同的计算机，相同的时间来做实时性分析是不可能的，因此本文利用浮点数来估计方案的计算复杂度，一个加号和一个乘号都叫作一个浮点，加号和乘号的总数作为总的浮点数，即浮点数的总和代表着计算复杂度。本文对 MTN 模型(CG-MTN)和精简 MTN 模型(WE-CG-MTN)的浮点数进行了对比，说明精简 MTN 模型浮点数远远小于 MTN 模型。同时在以往研究中分析了 MTN 模型浮点数远远小于神经网络模型。这些说明本文所提方案有较低的计算复杂度，拥有较好的实时性能。

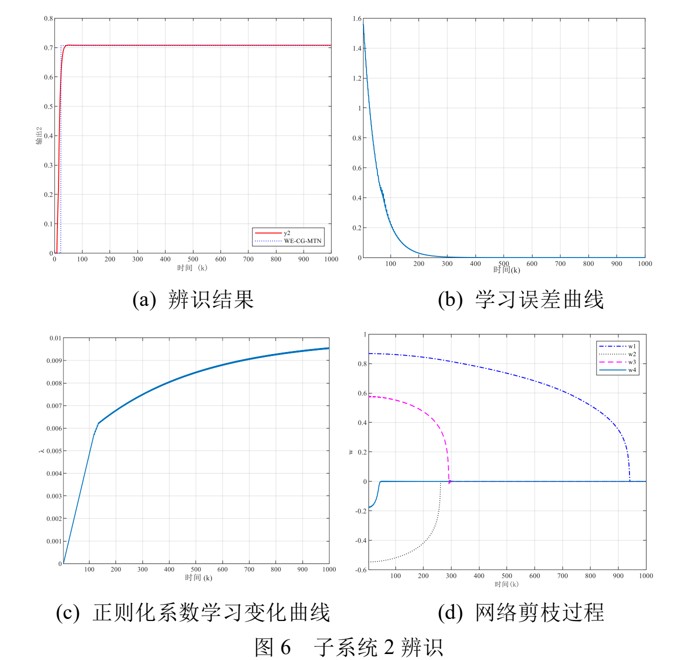


(a) 辨识结果　　(b) 学习误差曲线

(c) 正则化系数学习变化曲线　　(d) 网络剪枝过程

图 5　子系统 1 辨识

Fig. 5　Identification of sub-system 1



(a) 辨识结果　　(b) 学习误差曲线

(c) 正则化系数学习变化曲线　　(d) 网络剪枝过程

图 6　子系统 2 辨识

Fig. 6　Identification of sub-system 2

4　结束语

针对 MIMO 非线性时滞系统提出了基于 MTN 的辨识方案。

a) 利用 MTN 的强大逼近性，MTN 作为辨识模型。

b) WE-CG 算法作为 MTN 辨识模型的学习算法；WE 算

法可以有效精简 MTN 辨识模型结构，从而降低算法复杂度，提高模型实时性能。

实验结果表明，本文所提辨识方案够准确地对 MIMO 非线性时滞系统进行辨识。同时，相比传统的 MTN 辨识方案，所提辨识方案结构更精简，具有更低的算法复杂度，实时性更好，为工程实践和硬件实施提供方便。

虽然所提辨识方法展现出良好的性能，但仍需进一步的研究，比如：a)研究其他的精简辨识方案；b)研究多入多出非线性时滞系统控制问题；c)研究该方案应用到其他工程领域问题。

参考文献：

[1] Liu X, Yang X Q, Zhu P B, *et al.* Robust identification of nonlinear time-delay system in state-space form [J]. Journal of the Franklin Institute, 2019, 356 (16): 9953-9971.

[2] 侯明冬, 王印松. 一类非线性大滞后系统自适应积分滑模控制 [J]. 控制理论与应用, 2019, 36 (7): 1182-1188. (Hou Mingdong, Wang Yinsong. An adaptive integral sliding mode control for a class of nonlinear large-lag systems [J]. Control Theory & Applications, 2019, 36 (7): 1182-1188.)

[3] Zhu Kun, Yu Chengpu, Wan Yiming. Recursive least squares identification with variable-direction forgetting via oblique projection decomposition [J]. IEEE/CAA Journal of Automatica Sinica, 2022, 9 (3): 547-555.

[4] Ahmed S. Step response-based identification of fractional order time delay models [J]. Circuits Syst Signal Process, 2020, 39: 3858-3874.

[5] 陈茹雯, 湛时时. 基于非线性自回归时序模型的振动系统辨识 [J]. 计算机应用研究, 2016, 33 (10): 3021-3025. (Chen Ruwen, Zhan Shishi. Identification based on GNAR model for vibration system [J]. Application Research of Computers, 2016, 33 (10): 3021-3025.)

[6] Lu Lu, Yu Yi, Yang Xiaomin, *et al.* Time delay Chebyshev functional link artificial neural network [J]. Neurocomputing, 2019, 329: 153-164.

[7] de Paula N C G, Marques F D. Multi-variable Volterra kernels identification using time-delay neural networks: application to unsteady aerodynamic loading [J]. Nonlinear Dynamics, 2019, 97: 767-780.

[8] Ren Xuemei, Rad A B. Identification of nonlinear systems with unknown time delay based on time-delay neural networks [J]. IEEE Transactions on Neural Networks, 2007, 18 (5): 1536-1541.

[9] Xu Jun, Tao Qinghua, Li Zhen, *et al.* Efficient hinging hyperplanes neural network and its application in nonlinear system identification [J]. Automatica, 2020, 116: 108906.

[10] Wu Xiaolong, Han Honggui, Liu Zheng, *et al.* Data-knowledge-based fuzzy neural network for nonlinear system identification [J]. IEEE Transactions on Fuzzy Systems, 2020, 28 (9): 2209-2221.

[11] Rummelhart D E, Hinton G E, Williams R. J. Learning representations by back propagating errors [J]. Nature, 1986, 323: 533-536.

[12] Batselier K, Chen Zhongming, Wong N. A Tensor Network Kalman filter with an application in recursive MIMO Volterra system identification [J]. Automatica, 2017, 84: 17-25.

[13] Chee E, Wang Xiaonan. Generalized system identification for nonlinear MPC of highly nonlinear MIMO systems [J]. IFAC-PapersOnLine, 2021, 54 (3): 366-371.

[14] Gray W S, Venkatesh G S, Espinosa L A D. Nonlinear system identification for multivariable control via discrete-time Chen-Fliess series [J]. Automatica, 2020, 119: 109085.

[15] 严洪森. 多维泰勒网优化控制 [EB/OL]. https://automation. seu. edu. cn/yhs/list. htm, 2021-07-06. (Yan Hongsen. MTN (multi-dimensional Taylor network) optimal control [EB/OL]. https://automation. seu. edu. cn/yhs/list. htm, 2021-07-06.)

[16] 严洪森. 等效论 (Equivology) 与多维泰勒网 [EB/OL]. https://automation. seu. edu. cn/yhs/list. htm, 2021-07-06. (Yan Hongsen. Equivology and MTN (multi-dimensional Taylor network) [EB/OL]. https://automation. seu. edu. cn/yhs/list. htm, 2021-07-06.)

[17] 周博, 严洪森. 基于小波和多维泰勒网动力学模型的金融时间序列预测 [J]. 系统工程理论与实践, 2013, 33 (10) , 2654-2662. (Zhou Bo, Yan Hongsen. Financial time series forecasting based on wavelet and multi-dimensional Taylor network dynamics model [J]. Systems Engineering-Theory & Practice, 2013, 33 (10): 2654-2662.)

[18] 林屹, 严洪森, 周博. 基于多维泰勒网的非线性时间序列预测方法及其应用 [J]. 控制与决策, 2014, 29 (5): 795-801. (Lin Yi, Yan Hongsen, Zhou Bo. Nonlinear time series prediction method based on multi-dimensional Taylor network and its applications [J]. Control and Decision, 2014, 29 (5): 795-801.)

[19] Li Chenlong, Yan Hongsen. Identification of nonlinear time-delay system using multi-dimensional Taylor network model [A]. In: The 8th International Conference on Manipulation, Manufacturing and Measurement on the Nanoscale (IEEE 3M-NANO 2018) [C]. Hangzhou, China, 2018: 87-90.

[20] 李晨龙, 严洪森. 基于多维泰勒网的超前 d 步预测模型 [J]. 控制与决策, 2021, 36 (2): 345-354. (Li Chenlong, Yan Hongsen. d-step-ahead predictive model based on multi-dimensional Taylor network [J]. Control and Decision, 2021, 36 (2): 345-354.)

[21] Li Chenlong, Yan Hongsen, Zhang Jiaojun. Multi-dimensional Taylor network adaptive predictive control for SISO nonlinear systems with input time-delay [J]. Transactions of the Institute of Measurement and Control, 2021, DOI: 10. 1177/01423312211040294.

[22] Zhang Jiaojun, Yan Hongsen. MTN optimal control of MIMO non-affine nonlinear time-varying discrete systems for tracking only by output feedback [J]. Journal of the Franklin Institute, 2019, 356 (8): 4304-4334.

[23] Li Chenlong, Yuan Changshun, Ma Xiaoshaung, *et al.* Integrated fault detection for nonlinear process monitoring based on Multi-dimensional Taylor Network [J]. Assembly Automation, 2022, https://doi. org/10. 1108/AA-06-2021-0076.

[24] Han Yuqun, Yan Hongsen. Adaptive multi-dimensional Taylor network tracking control for SISO uncertain stochastic nonlinear systems [J]. IET Control Theory and Applications, 2018, 12 (8): 1107-1115.

[25] Klambauer G. Mathematical analysis [M]. New York: Marcel Dekker INC, 1975: 236-237.

[26] 丁锋. 系统辨识算法的复杂性、收敛性及计算效率研究 [J]. 控制与决策, 2016, 31 (10): 1729-1741. (Ding Feng. Complexity, convergence and computational efficiency for system identification algorithms [J]. Control and Decision, 2016, 31 (10): 1729-1741.)

[27] Golub G H, Van Loan C F. Matrix computations (Vol. 3) [M]. Baltimore: JHU Press, 2012.

[28] 魏海坤. 神经网络结构设计的理论与方法 [M]. 北京：国防工业出版社, 2005. (Wei Haikun. Theory and method of neural network structure design [M]. National Defence Industry Press. Beijing, China, 2005.)

[29] 魏海坤, 丁维明, 宋文忠, 等. RBF 网的动态设计方法 [J]. 控制理论与应用, 2002, 19 (5): 673-680. (Wei Haikun, Ding Weiming, Song Wenzhong, *et al.* Dynamic method for designing RBF neural networks [J]. Control Theory and Applications, 2002, 19 (5): 673-680.)

[30] Weigend A S, Rumelhart D E, Huberman B A. Generalization by weight-elimination with application to forcasting. In: Advances in Neural Information Processing Systems [C]. San Mateo, 1991, 3: 857-882.

[31] Zerari N, Chemachema, M. Robust adaptive neural network prescribed performance control for uncertain CSTR system with input nonlinearities

and external disturbance [J]. Neural Computing Applications. 2020, 32: 10541-10554.

[32] Wu Qun, Du Wenli, Nagy Z. Steady-state target calculation integrating economic optimization for constrained model predictive control [J]. Computers & Chemical Engineering, 2021, 145: 107145.

[33] Yang Xiong, Wei Qinglai. Adaptive critic designs for optimal event-driven control of a CSTR system [J]. IEEE Transactions on Industrial Informatics, 2021, 17 (1): 484-493.

[34] Du Jing Jing, Johansen T A. A gap metric based weighting method for multimodel predictive control of MIMO nonlinear systems [J]. Journal of Process Control, 2014, 24: 1346-1357.

[35] Tan Yonghong, Cauwenberghe A V. Neural-network-based d-step-ahead predictors for nonlinear systems with time delay [J]. Engineering Applications of Artificial Intelligence, 1999, 12: 21-35.